IS 7600 (1975): Analysis of variance [MSD 3: Statistical Methods for Quality and Reliability]

BLANK PAGE

Indian Standard

(Reaffirmed 2001)

ANALYSIS OF VARIANCE

(Third Reprint MAY 1996)

UDC 519.241.6

BUREAU OF INDIAN STANDARDS
MANAK BHAVAN, 9 BAHADUR SHAH ZAFAR MARG
NEW DELHI 110002

Gr 5

August 1975

Indian Standard
ANALYSIS OF VARIANCE

Quality Control and Industrial Statistics Sectional Committee, EC 3

(*Continued on page* 2)

(*Continued from page* 1)

Members	*Representing*
SHRI S. N. VOHRA	Directorate General of Supplies and Disposals, New Delhi
SHRI Y. K. BHAT, Deputy Director (Stat) (*Secretary*)	Director General, ISI (*Ex-officio Member*)

Industrial Statistics Subcommittee, EC 3 : 7

Convener	
DR P. K. BOSE	University of Calcutta, Calcutta; *and* Indian Institute of Social Welfare and Business Management, Calcutta
Members	
DIRECTOR	Institute of Agricultural Research Statistics (ICAR), New Delhi
DR M. HOLLA (*Alternate*)	
DR A. K. GAYEN	Indian Institute of Technology, Kharagpur
SHRI S. K. GUPTA	Central Statistical Organization, New Delhi
SHRI S. B. PANDEY	Imperial Chemical Industries (India) Private Ltd, Calcutta
SHRI B. K. SARKAR	Indian Statistical Institute, Calcutta
SHRI D. R. SEN	Delhi Cloth & General Mills Co Ltd, Delhi
DR (KUMARI) N. S. SHAKUNTALA	Defence Research and Development Organization (Ministry of Defence), New Delhi
SHRI B. SITARAMAN	Army Statistical Organization (Ministry of Defence), New Delhi
SHRI P. N. KAPOOR (*Alternate*)	

Indian Standard
ANALYSIS OF VARIANCE

0. FOREWORD

0.1 This Indian Standard was adopted by the Indian Standards Institution on 10 February 1975, after the draft finalized by the Quality Control and Industrial Statistics Sectional Committee had been approved by the Executive Committee.

0.2 The technique of analysis of variance is an indispensable tool in the scientific and industrial research for the analysis of experimental data involving quantitative measurements and is particularly helpful when several independent sources of variation are present in the data such as the results obtained at different temperatures, duplicate determinations of the same material made by several analysts, measurements classified according to several sources of supply of raw material obtained from different vendors, etc.

0.3 It is well-known that the observations obtained by repetitive experiments vary among themselves. The source of variation in the data may be due to various causes, assignable or chance. Using the analysis of variance techniques it is possible to estimate how much of the total variation in a set of data can be attributed to one or more assignable causes of variation, the remainder which is not attributable to any assignable causes of variation being classed as due to chance causes which produces the residual or error variation.

0.4 This standard is a sequel to the 'Indian Standard on Statistical Tests of Significance' (IS : 6200-1971). To compare the means of two groups of observations and to assess whether the difference between them can be reasonably ascribed to chance, the *t*-test is used. When the comparison is to be made among the means of more than two groups of observations, resort to the analysis of variance technique is made. In fact, *t*-test is a particular case of analysis of variance.

0.5 In reporting the result of a test or analysis, if the final value, observed or calculated, is to be rounded off, it shall be done in accordance with IS : 2-1960*.

1. SCOPE

1.1 This standard intends to give a brief outline of the general treatment of the analysis of variance technique with respect to some of the designs

*Rules for rounding off numerical values (*revised*).

which are more frequently used in industrial experimentation. The techniques have been illustrated with examples wherein the necessary computational details have also been given.

2. TERMINOLOGY

2.0 For the purpose of this standard, the following definitions shall apply.

2.1 Standard Deviation — The square root of the quotient obtained by dividing the sum of squares of deviations of the observations from their mean by one less than the number of observations in the sample.

2.2 Variance — Square of standard deviation.

2.3 Degrees of Freedom (D. F.) — The number of independent component values which are necessary to determine a statistic.

2.4 Null Hypothesis — The hypothesis (or assumption) of the equivalence (or no difference) among the effects of methods so that the samples emanate from the same lot.

2.5 Level of Significance — The probability (or risk) of rejecting the null hypothesis when it is true. Conventionally, it is taken to be 5 percent or 1 percent.

2.6 Corrected Sum of Squares — The total of the squares of the deviations of the observations from their mean.

3. SOME BASIC CONCEPTS

3.1 Mathematical Model

3.1.1 Before the application of the analysis of variance techniques to any experimental data, it is fundamental to have some knowledge of the mathematical model holding good for the particular investigation under consideration. Basically there are three models, namely, fixed effects model, random effects model and mixed effects model (*see also* Appendix A).

3.1.2 In the fixed effects model, some of the assignable causes of variation in the experiment are deliberately chosen so that the results of the analysis are not amenable for generalization in that direction. In other words, when the effects are unknown constants (parameter) the model is called fixed effects model or model I. For example, in an inter-laboratory investigation on the checking of the precision of tensile testing machine, four laboratories possessing a particular brand of the machine may be intentionally chosen so that the results of the analysis will apply only to these four laboratories and any conclusion derived will not be applicable to all the laboratories in general. On the other hand, if the tensile testing machines of a large number of laboratories are to be investigated and due to limitations of facilities, the four laboratories chosen are a random sample of all the laboratories, then the mathematical model chosen is that of a random effects model or model II. The mixed effects model, as the name itself indicates, is the combination of the earlier two models where some of the effects are of the fixed nature, the remaining being of random nature. The random effects model has extensive applications in most of the industrial experimentation.

3.2 Additive Nature of Sum of Squares

3.2.1 As the name implies the technique of analysis of variance consists in separating the total variance into parts, each part measuring the variability attributable to some specific source. For example in an inter-laboratory testing there may be variability among laboratories and variability within the laboratory. The latter, in turn, may be composed of a number of components of differing magnitude and importance like the variability among different analysts, among days, or among determinations made on the same day. Using the additive property of the variance the total variation between members of a set of observations, classified according to one or more criteria, can be broken up into components, attributable to different criteria of classification which are of experimental interest or importance. By testing the significance of these components it is possible to determine which of the criteria are associated with a significant proportion of the overall variability in the averages.

3.2.2 Planned experiments have proved to be very powerful and economical in investigating the influence of various factors contributing to the total variance in the measurable characteristics of the product. In these, measurements are taken on a sample of units, which is so constituted as to ensure the simultaneous randomization in experimental error and variation due to changes in treatments. As long as they are clearly defined and reproducible these treatments may represent different materials or temperatures or processes or any variation in operating conditions.

3.3 Orthogonality of Designs

3.3.1 For the easy statistical analysis of the data resulting from any investigation, a desirable feature of the design adopted is its orthogonality. Orthogonality ensures that the different classes of effects shall be capable of direct and separate estimation without any entanglement.

3.3.2 For example, in an experiment to observe the effect of varying two factors, say temperature and pressure, wherein the temperature is measured at four levels and pressure at three levels, there will be 12 experimental conditions generated by taking each level of temperature with each level of pressure. The comparison between the average of three results at any two temperatures will then be purely a measure of the effect of temperature (and *vice versa*) since the averages will have been taken over the same set of pressures. The two factors are said to be mutually orthogonal.

Now by accident or design suppose that the combination corresponding to the highest temperature and pressure is missing from the experiment, then the difference between the averages of the observations at the lowest and the highest temperatures could not be purely a measure of the effect of temperature because the first average is based on observations at three levels of pressure whereas the second observation is based on the average at only two levels of pressure and hence the difference between the two averages will be, to an unknown extent, influenced by the effect of pressure. The experiment in this case is non-orthogonal.

3.4 The details of the application of the technique of analysis of variance will vary with the number of independent causes of variation. It is possible to classify the data with respect to each independent source of variation and the complete classification is a necessary first step to the application of analysis of variance. The following sections describe the procedure for analysis of variance under different categories which are commonly met in practice.

4. ANALYSIS OF VARIANCE — SINGLE FACTOR OR ONE-WAY CLASSIFICATION OF DATA

4.1 Replicate Determinations

4.1.1 Data will frequently be encountered where classification is based on one factor only, for example analysts or temperatures or batches of material. The data may consist of (a) replicate determinations of the same material made by several analysts in the same laboratory or (b) measurements obtained at different temperatures or (c) measurable characteristics of some material obtained in different lots or shipment from the same supplier, etc.

4.1.2 In all these, there will be variation within replicates (unassignable variation in the system) which is a measure of the precision. There will also be a variation in means of results obtained under different conditions. It is due to the differences among analysts or differences in temperatures or lot-to-lot variation of the shipment. Analysis of variance helps to separate these effects and to determine whether there is any significant difference between operators or temperatures or shipments, as the case may be.

4.1.3 Usually, in the interest of experimental efficiency and simplicity of analysis, it is desirable to have the same number of replicate determinations for each class. But owing to a lack of design or to the loss of part of data or natural grouping of experimental material or perhaps due to deliberate placing of emphasis on certain effects, the number of observations in various classes may be unequal. The following two examples illustrate the method of analysis when there are (a) unequal number of observations in each class, and (b) equal number of observations in each class.

4.1.4 *Example* 1 — The following data represent the warpway breaking strength of Type II Indian hessian [*see* IS : 2818 (Part II)-1971*] measured

*Specification for Indian hessian: Part II 305 and 229 g/m² at 16 percent contract regain (*first revision*).

in units of kg. Products of three different mills A, B and C were tested and the test results are reproduced below:

Mill	*Warpway Breaking Strength* (kg)	*No. of Observations*	*Total*	*Mean*
A	87, 96, 99, 94, 91	5	467	93·4
B	93, 103, 90, 93, 99, 88, 100, 91	8	757	94·6
C	98, 88, 84	3	270	90·0
	Total	16	1 494	

The various calculations needed for forming the required analysis of variance are as follows:

a) Uncorrected total sum of squares : $(87)^2+(96)^2+\ldots.(84)^2=139\,940$

b) Uncorrected sum of squares between mills : $\frac{(467)^2}{5}+\frac{(757)^2}{8}+\frac{(270)^2}{3}$
$=139\,548{\cdot}93$

c) Correction factor : $\frac{(1\,494)^2}{16}=139\,502{\cdot}25$

d) Total sum of squares : $(a)-(c)=437{\cdot}75$

e) Sum of squares between mills : $(b)-(c)=46{\cdot}68$

f) Sum of squares within mills : $(d)-(e)=391{\cdot}07$

The degrees of freedom corresponding to the various entries of the analysis of variance table are obtained as follows:

Degrees of freedom for total sum of squares = Total number of observations $-1=16-1$
$=15$

Degrees of freedom for between mills = Total number of mills -1
$=3-1=2$

Degrees of freedom for within mills = Total degrees of freedom $-$ degrees of freedom for between mills $=15-2=13$

Analysis of Variance Table

Source of Variation	*Degrees of Freedom*	*Sum of Squares*	*Mean Square*	*F Ratio*
Between mills	2	46·68	23·34	$\frac{23{\cdot}34}{30{\cdot}08}=0{\cdot}78$
Within mills	13	391·07	30·08	
Total	15	437·75		

To test the hypothesis that there is no appreciable variation between the average breaking strength of the products of the different mills the variance ratio is formed as $F=\frac{23{\cdot}34}{30{\cdot}08}=0{\cdot}78$. Since this is less than 3·81 which is the tabulated value of F for 2 and 13 degrees of freedom at 5 percent level, there is insufficient evidence to reject the hypothesis. Therefore, it may be concluded that there is no appreciable variation in the breaking strength of the products of the three mills.

4.1.5 *Example* 2 — The following data are from 5 replicate runs on the time of passage of aluminium powder through 6 different test units. It is endeavoured to test whether there is any significant difference between the 6 test units when compared with the variation within units:

Run	*Unit*→	1	2	3	4	5	6
1		52·9	54·0	52·6	50·5	54·6	54·0
2		52·3	53·8	53·2	50·8	54·6	53·3
3		52·2	53·8	53·4	50·7	54·4	53·7
4		52·5	53·6	53·4	50·8	54·4	53·5
5		52·7	53·6	53·0	50·5	54·4	53·7

The data can be coded by subtracting 50 from each value and multiplying the remainder by 10 to remove the decimal. Coding by adding or subtracting a constant has no effect on the calculation of the corrected sum of squares or the variance. Coding by multiplying by a constant will change the resulting variance by the square of this constant. However, since the mean squares are tested by a ratio of two calculations, the coding factor will cancel. If ultimately it is intended to use the calculated variances, they will have to be decoded by dividing by 100.

The coded data and the analysis of variance are as follows:

Run	*Unit*→	1	2	3	4	5	6	
1		29	40	26	5	46	40	
2		23	38	32	8	46	33	
3		22	38	34	7	44	37	
4		25	36	34	8	44	35	
5		27	36	30	5	44	37	
	Total	126	188	156	33	224	182	909
	Mean	25·2	37·6	31·2	6·6	44·8	36·4	

The various calculations are obtained as below:

a) Uncorrected total sum of squares: $(29)^2+(23)^2+\ldots+(35)^2+(37)^2 = 32\ 119$

b) Uncorrected sum of squares between units: $\frac{1}{5}[(126)^2+(188)^2+\ldots(182)^2] = 31\ 989$

c) Correction factor: $\frac{(909)^2}{30} = 27\ 542{\cdot}7$

Hence the corrected sum of squares are obtained as:

d) Total sum of squares: (a)−(c) $= 32\ 119 - 27\ 542{\cdot}7 = 4\ 576{\cdot}3$

e) Sum of squares between units: (b)−(c) $= 31\ 989 - 27\ 542{\cdot}7 = 4\ 446{\cdot}3$

f) Sum of squares within units: (d) − (e) = 4 576·3 − 4 446·3 = 130·0

The degrees of freedom for the various entries of the analysis of variance table are obtained as follows:

Total sum of squares	:	30 − 1 = 29
Between units	:	6 − 1 = 5
Within units	:	29 − 5 = 24

ANALYSIS OF VARIANCE TABLE

Source of Variation	*D.F.*	*Sum of Squares*	*Mean Square*	*F Ratio*
Between units	5	4 446·3	889·26	$\frac{889{\cdot}26}{5{\cdot}42} = 164{\cdot}1$**
Within units	24	130·0	5·42	
Total	29	4 576·3		

**Highly significant.

Tabulated F value for 5 and 24 degrees of freedom at 0·01 level is only 3·90. The F value as obtained from the data is highly significant. Hence there is a strong evidence of a factor between units which causes a variation in the results greater than that which can be accounted for by variation within units.

It may now be desirable to compare all possible pairs of means to find out which of the means or how many of them differ significantly from the others to cause the overall variation among means to be significant. For this purpose use is made of t-test for difference between means, taken two at a time. With n means, there are $\frac{n(n-1)}{2}$ comparisons (number of combinations of n things taken two at a time). In the present example there will be $\frac{6 \times 5}{2} = 15$ comparisons.

The residual variance gives a measure of the precision of a single measurement. This measure is identical with the pooled estimate of the standard deviation. In the present example it is obtained as $\sqrt{5{\cdot}42} = 2{\cdot}327$. Standard deviation of the mean of five measurements is $\frac{1}{\sqrt{5}}$ times the standard deviation of single measurement and is equal to $\frac{2{\cdot}327}{\sqrt{5}} = 1{\cdot}040\,6$.

The critical difference, useful for finding out whether any two means are significantly different or not, is obtained as $t_{24} \times \sqrt{2} \times 1{\cdot}040\,6$ where t_{24} is the tabulated value of the t distribution for 24 degree of freedom at the 5 percent level, which is obtained as 2·064. The critical difference is thus obtained as 3·04. It may, however, be noted that the use of critical difference is valid only when the F ratio in the analysis of variance table is found to be significant. After obtaining the critical difference the various means are arranged either in the ascending or the descending order of magnitude as follows :

Unit No.	4	1	3	6	2	5
Means	6·6	25·2	31·2	36·4	37·6	44·8

Any two means not under scored by the same line are significantly different. Thus in the present example only the means between unit 6 and unit 2 are not different. Hence it may be concluded that all other units give results which are significantly different from one another.

4.2 Single Classification with Subgrouping (Nesting)

4.2.1 Often the principal classification in the data can be divided into sub-classifications which do not cut across the main classes. Each sample may be composed of sub-samples and these, in turn, may consist of sub-sub-samples. The repeated sampling and sub-sampling give rise to nested sampling or hierarchical classification. For example if we draw 10 bales of wool at random from a shipment and take three cores of wool from each bale, the cores are said to be nested within bales. Again several analysts might draw two, three or four specimens (not necessarily the same number for each analyst) from a batch of material and run several replicate analyses on each specimen. The replicate analyses give an estimate of the error variance. The specimens run by the same analyst give a measure of variation within the batch plus the error. The differences between the results of the several analysts include not only the variation between analysts but also the variation between specimens and the error variation. Same situation holds good if several test pieces are cut from each of several rolls of fabric made wholly from several different machines from each of several plants. As long as there is no relation between the corresponding members of the different groupings, the analysis of variance by subgroups (or nesting, as it is sometimes called), applies. Where there is relationship between the members of the subgroups, so that the corresponding division of each subgroup can be considered as a separate class, we have an analysis of variance for more than one main classification. The general arrangement of data for single classification with several hierarchies of subgrouping is illustrated by the following example.

4.2.2 *Example* 3 — A series of trials is made by three operators to locate a source of variation in a chemical analysis. The procedure consists in taking a specimen, treating it in a combustion-tube furnace and performing the chemical analysis. In the test, three operators each took two specimens and made three combustion trials on each specimen and titrated each trial in duplicate. The single letter A represents the operator factor, the double

DATA FOR EXAMPLE 3

A		1						2						3					
B(*A*)		11			12			21			22			31			32		
C(*AB*)		111	112	113	121	122	123	211	212	213	221	222	223	311	312	313	321	322	323
Duplicate results		156	151	154	148	154	147	125	94	98	118	112	98	184	172	181	172	181	175
		154	154	160	150	157	149	125	95	102	124	117	110	184	186	191	176	184	177
Sub Totals	i)	310	305	314	298	311	296	250	189	200	242	229	208	368	358	372	348	365	352
	ii)	929			905			639			679			1 098			1 065		
	iii)	1 834						1 318						2 163					
Grand Total		5 315																	

letter $B(A)$ indicates the first or second specimen taken by the operator, and the triple letter $C(AB)$ indicates the combustion trial on each specimen by each operator. The results are shown in duplicate.

Uncorrected sums of squares are obtained as follows:

a) Total sum of squares: $(156)^2+(151)^2+\ldots+(184)^2+(177)^2 = 817\,085$

b) Combustion within operators and specimens $C(AB)$:

$$\frac{(310)^2+(305)^2+\ldots+(365)^2+(352)^2}{2} = 816\,778$$

c) Specimen within operator $B(A)$:

$$\frac{(929)^2+(905)^2+\ldots+(1\,098)^2+(1\,065)^2}{6} = 815\,209$$

d) Between operator:

$$\frac{(1\,834)^2+(1\,318)^2+(2\,163)^2}{12} = 814\,937$$

Correction factor: $\frac{(5\,315)^2}{36} = 784\,700$

ANALYSIS OF VARIANCE TABLE

Source of Variation	*D.F.*	*Sum of Squares*	*Mean Square*	*F Ratio*
Operators A	2	814 937−784 700=30 237	15 168	$\frac{15\,168}{131} = 116$**
Specimen within operators $B(A)$	3	815 209−814 937= 272	91	
Combustion within operators & specimens $C(AB)$	12	816 778−815 209= 1 569	131	$\frac{131}{17} = 7{\cdot}7$**
Replicates, error	18	(By subtraction) 307	17	
Total	35	817 085−784 700=32 385		

**Highly significant.

From the analysis of variance table it may be concluded that the largest source of variation in results is between operators. There is no evidence of variation between specimens. There is a definite source of variation in combustion step in analysis.

5. ANALYSIS OF VARIANCE — TWO-WAY CLASSIFICATION OF DATA

5.1 If a series of experiments is run at different temperatures and different pressures or if material from several sources of supply is tested under a variety of conditions or if a group of operators makes a series of runs on a number of pilot plants, two-way classification of data is obtained. In such a case, one of the characteristics can be represented along the rows and the other along the columns. Further, each of the cells formed by the two-way classification can have either one or more than one observation.

5.2 Two-Way Classification with One Observation in Each Cell

5.2.1 The analysis of variance for this arrangement of data is similar to that for the single factor arrangement. The column-factor effect is calculated from the squares of column totals and the row factor effect, from the squares of row-totals. The following example illustrates the method of analysis.

5.2.2 *Example* 4 — Six samples of dextrose monohydrate were analysed in each of the seven laboratories for copper content (measured as ppm). The data obtained by the investigation is given below:

Sample / *Lab*	1	2	3	4	5	6	*Total*
1	0·3	0·2	0·1	0·7	0·5	0·4	2·2
2	0·9	0·9	0·3	0·3	0·3	0·8	3·5
3	0·8	1·9	0·6	0·4	0·4	1·1	5·2
4	0·6	0·6	0·2	0·3	0·3	0·3	2·3
5	0·5	0·2	0·2	0·2	0·5	0·5	2·1
6	0·4	0·4	0·5	0·4	0·6	0·6	2·9
7	0·5	0·6	0·1	1·0	1·2	1·2	4·6
Total	4·0	4·8	2·0	3·3	3·8	4·9	22·8

The uncorrected sum of squares are calculated as follows:

a) Total sum of squares : $(0{\cdot}3)^2+(0{\cdot}2)^2+\ldots+(1{\cdot}2)^2=17{\cdot}56$

b) Between laboratories sum of squares : $\frac{(2{\cdot}2)^2}{6}+\frac{(3{\cdot}5)^2}{6}+\ldots+\frac{(4{\cdot}6)^2}{6}=13{\cdot}90$

c) Between samples sum of squares : $\frac{(4{\cdot}0)^2}{7}+\frac{(4{\cdot}8)^2}{7}+\ldots+\frac{(4{\cdot}9)^2}{7}=13{\cdot}20$

Correction factor : $\frac{(22{\cdot}8)^2}{42}=12{\cdot}38$

Analysis of variance table is formed as follows:

ANALYSIS OF VARIANCE TABLE

Source of Variation	*D.F.*	*Sum of Squares*	*Mean Square*	*F Ratio*
Between rows (laboratories)	6	13·90 − 12·38 = 1·52	0·25	$\frac{0{\cdot}25}{0{\cdot}09}=2{\cdot}78$
Between columns (samples)	5	13·20 − 12·38 = 0·82	0·16	$\frac{0{\cdot}16}{0{\cdot}09}=1{\cdot}78$
Error	30	(by subtraction) 2·84	0·09	
Total	41	17·56 − 12·38 = 5·18		

From the analysis of variance table, it can be concluded that there is no significant difference between either the laboratories means or the sample means.

5.3 Two-Way Classification with Multiple Observations in Each Cell — In this type of classification more than one observation is obtained in each cell formed by the rows in the columns, that is, repeated measurements are made on different randomly selected individuals.

5.3.1 Replication facilitates fuller analysis of the data and the precision of the experiment increases with replication.

5.3.2 In the two-way classification with more than one observation per cell, besides the variation due to row and column effects there would also be an interaction effect which is the result of different row column combinations. As all the different combinations from the groups of the two factors of classification play their part in the experiment, they also contribute to the total variability. This interaction becomes one of the sources of variability which must be taken into account in the analysis of variance. Supposing the criteria of classification are different varieties of material subjected to different treatments. In such a situation the same variety can be differently affected by different treatments and the same treatment can show different effects with different varieties. This combination effect is the interaction. The following example illustrates the method of analysis.

5.3.3 *Example* 5 — For studying the accuracy of water meters for continuous rate of flow, 5 water meters were tested by each of the two operators *A* and *B*. Each operator made five repeat observations on each meter. The observations in terms of the percentage accuracy of the meters are given below :

Operator		*Water Meter Number*					*Total*
		1	2	3	4	5	
A		+ 0·5	− 3	− 2·5	2	− 3	
		− 1	− 2·5	− 3·5	2	− 3·5	
		− 1	− 3	− 1·5	2	− 3	
		− 1	− 2	− 2	2	− 3	
		− 1	− 3	− 2·5	2	− 3	
	Cell total	− 3·5	−13·5	−12·0	10·0	−15·5	−34·5
B		+ 0·5	− 3	− 3	0	− 3	
		− 3	− 2	− 2	0	− 3	
		− 2	− 2·5	− 2	0	− 4	
		− 1	− 3·0	− 2	0	− 3·5	
		− 1·5	− 2	− 1	2	− 3	
	Cell total	− 7·0	−12·5	−10·0	2·0	−16·5	−44·0
	Total	−10·5	−26·0	−22·0	12·0	−32·0	−78·5

From the above table the various sums of squares are obtained as follows :

a) Uncorrected total sum of squares $=(0\cdot5)^2+(-1)^2+\ldots.(-3)^2$
$=270\cdot75$

Correction factor (CF) $= \frac{(-78\cdot5)^2}{50} = 123\cdot24$

Hence the corrected total sum of squares $=270\cdot75-123\cdot24=147\cdot51$

b) Between meters sum of squares $= \frac{(-10\cdot5)^2}{10} + \ldots. \frac{(-32\cdot0)^2}{10} - \text{CF}$
$= 243\cdot82-123\cdot24$
$= 120\cdot58$

c) Between operators sum squares $= \frac{(-34\cdot5)^2}{25} + \frac{(-44\cdot0)^2}{25} - \text{CF}$
$= 125\cdot05-123\cdot24=1\cdot81$

d) Interaction sum of squares
$= \frac{(-3\cdot5)^2}{5} + \ldots. \frac{(16\cdot5)^2}{5} - \text{CF}$ — between meters sum of squares — between operators sum of squares
$= 252\cdot05-123\cdot24-120\cdot58-1\cdot81$
$= 6\cdot42$

The analysis of variance table is then formed as follows :

Source of Variation	*Degree of Freedom*	*Sum of Squares*	*Mean Square*	*F Ratio*
Between meters (M)	4	120·58	30·14	$\frac{30\cdot14}{1\cdot60} = 18\cdot84$**
Between operator (O)	1	1·81	1·81	$\frac{1\cdot81}{1\cdot60} = 1\cdot13$
Interaction ($M \times O$)	4	6·42	1·60	$\frac{1\cdot60}{0\cdot47} = 3\cdot40$*
Error	40	18·70	0·47	
Total	49	147·51		

**Highly significant.
*Significant.

When the interaction is tested against the error it is found to be highly significant thereby indicating the presence of interaction. It may hence be interpreted that different water meters behave differently with the change of operator. In view of this finding, there is an urgent need for the procedure for testing of the water meters to be standardized. Since the interaction is significant, under the random effects model chosen for the experiment, meters and operators are to be tested against the interaction. This testing reveals that between operators variation is not significant whereas between meters variation is highly significant. The latter finding is perhaps to be expected, since the meters had originated from different manufacturers.

5.4 Two-Way Classification with Sub-grouping (Nesting and Replication)

5.4.1 Sub-grouping or nesting of data within a main classification can occur with data that are collected under two main classifications. For example, if several factories are producing the same product on a batch basis and each makes several batches at two or more conditions then a situation illustrated by nesting is obtained.

5.4.2 The batches from each factory are not related to batches from other factories and for the purpose of analysis of variance they are simply sub-groups of the factory classification. Any variation between factories would include the variation between batches within factories. The condition factor is an independent classification and its effect on the variation of results is reflected in the difference between factories, that is to say, only in so far as there is interaction between the conditions and factory factors or the condition and batch factors.

5.4.3 Sub-grouping can exist under either factor or both the factors in the same set of data. Secondary sub-grouping can exist within the first sub-groups. In fact, there can be any hierarchy of sub-groupings under both main classes of factors. An illustrative example is given wherein sub-grouping occurs under both the factors.

5.4.4 *Example* 6—For studying the effect of storage time and packing on the moisture content of corn flakes, two types of packings namely, polythene bags and polythene bags in cartons were chosen and the periods selected were 2, 4 and 6 months. Six different samples of corn flakes belonging to the same batch of manufacture were analysed in duplicate after storing them for 2, 4 and 6 months in the two types of packings. The resultant data is given below:

Storage Period	*Packing Type* 1			*Packing Type* 2		
	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3
2 months	6·17	4·40	4·22	4·66	3·98	6·22
	6·11	4·72	3·80	5·00	4·00	5·40
Cell Total	12·28	9·12	8·02	9·66	7·98	11·62
4 months	5·50	4·46	4·61	4·52	4·56	5·88
	5·48	4·85	4·25	4·43	3·17	4·83
Cell Total	10·98	9·31	8·86	8·95	7·73	10·71
6 months	6·10	5·80	6·79	6·89	7·12	6·55
	6·13	5·73	7·73	6·74	7·06	6·65
Cell Total	12·23	11·53	14·52	13·63	14·18	13·20

From the above data a sub-table of the following type is formed to assist in the computations.

SUB-TABLE

Storage Period	*Packing*		*Total*
	Type 1	Type 2	
2 months	29·42	29·26	58·68
4 months	29·15	27·39	56·54
6 months	38·28	41·01	79·29
Total	96·85	97·66	194·51

The various sum of squares are then obtained as follows:

a) Uncorrected total sum of squares $= (6{\cdot}17)^2 + \ldots (6{\cdot}65)^2 = 1\,095{\cdot}181\,9$

Correction factor (CF) $= \dfrac{(194{\cdot}51)^2}{36} = 1\,050{\cdot}948\,3$

Corrected total sum of squares $= 1\,095{\cdot}181\,9 - 1\,050{\cdot}948\,3 = 44{\cdot}233\,6$

b) Between packing sum of squares $= \dfrac{(96{\cdot}85)^2}{18} + \dfrac{(97{\cdot}66)^2}{18} - \text{CF}$

$= 1\,050{\cdot}966\,6 - 1\,050{\cdot}948\,3 = 0{\cdot}018\,3$

c) Between periods sum of squares $= \dfrac{(58{\cdot}68)^2}{12} + \dfrac{(56{\cdot}54)^2}{12} + \dfrac{(79{\cdot}29)^2}{12} - \text{CF}$

$= 1\,077{\cdot}251\,5 - 1\,050{\cdot}948\,3 = 26{\cdot}303\,2$

d) Interaction (packing $\times$ period) sum of squares

$= \dfrac{(29{\cdot}42)^2}{6} + \ldots \dfrac{(41{\cdot}01)^2}{6} - \text{CF} -$ between packing sum of squares

$-$ between period sum of squares

$= 1\,078{\cdot}132\,8 - 1\,050{\cdot}948\,3 - 0{\cdot}018\,3 - 26{\cdot}303\,2$

$= 0{\cdot}863\,0$

e) Between samples (within packing and period) sum of squares

$= \left[\dfrac{(12{\cdot}28)^2}{2} + \ldots \dfrac{(13{\cdot}20)^2}{2}\right] - \left[\dfrac{(29{\cdot}42)^2}{6} + \ldots \dfrac{(41{\cdot}01)^2}{6}\right]$

$= 1\,092{\cdot}521\,4 - 1\,078{\cdot}132\,8 = 14{\cdot}388\,6$

The analysis of variance table is then formed as follows:

From the analysis of variance table, between sample sum of squares (within packing and period) is tested against error and the *F* ratio so obtained is highly significant. Because of this fact both the main effects due to packing and period as also the interaction are tested against between sample (within packing and period). The testing reveals that the mean value of moisture content for between periods is highly significant, corroborating the general presumption that corn flakes gather moisture depending on the period of storage. The analysis also reveals that there is no significant difference in the moisture content of corn flakes stored in two different packings.

ANALYSIS OF VARIANCE TABLE

Source of Variation	*Degree of Freedom*	*Sum of Squares*	*Mean Square*	*F Ratio*
Between packing	1	0·018 3	0·018 3	$\frac{0·018\ 3}{1·199\ 0} = 0·002$
Between periods	2	26·303 2	13·151 6	$\frac{13·157\ 6}{1·199\ 0} = 10·97^{**}$
Packing period	2	0·863 0	0·431 5	$\frac{0·431\ 5}{1·199\ 0} = 0·36$
Between samples (within packing and period)	12	14·388 6	1·199 0	$\frac{1·199\ 0}{0·147\ 8} = 8·11^{**}$
Error	18	2·660 5 (by subtraction)	0·147 8	
Total	35	44·233 6		

**Highly significant.

5.5 Three-Way Classification — If there are three different sources of variation (factors of classification) which are likely to act independently of one another and contribute to the total variability of the data, then the total sum of squares has to be split up into three components due to these sources. The analysis of variance follows the same formulation as that of a two factor arrangement. When more than two factors are involved, all possible combinations of interactions may exist and a complete analysis of variance provides mean squares attributable to all the main factors and all individual interactions.

5.6 Factorial Experiments — When the effect of several variables on a product or process is of interest, it is possible to devise experiments where all of them may be studied simultaneously. For each variable, a number of categories or levels may be chosen for study. If an equal number of observations is made for all possible combinations of levels (one level from each variable), the experiment is called factorial. In a completely balanced experiment, each level of each factor is tested at all the levels of all the other factors so that the total number of observations required is the product of all the levels and all the factors. Thus in a factorial experiment to study the wear resistance of vulcanized rubber wherein five qualities of filler (factor *A*), three methods of pretreatment of the rubber (factor *B*) and four qualities of raw rubber (factor *C*) are involved, the total number of observations for a complete experiment turns out to be $5 \times 3 \times 4 = 60$. When more than two factors are involved, all the possible combinations of interactions may exist and a complete analysis of variance provides mean squares attributable to all individual interactions.

6. CLASSIFICATION OF HIGHER ORDER

6.1 There is no limit to the number of main factors like different treatments, materials, laboratories, temperatures, pressures, catalysts or concentrations, that may be examined in the same experiment. A suitable design of experiment is a pre-requisite in such studies.

APPENDIX A

(*Clause* 3.1)

FIXED, RANDOM AND MIXED MODELS

If x_{ijk} is the dimension of kth component produced on the ith machine on jth day in a plant, then it may be written that

$$x_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk}$$

where

μ = overall mean dimension,
a_i = effect due to ith machine,
b_j = effect due to jth day,
c_{ij} = effect due to interaction of ith machine and jth day, and
e_{ijk} = random effects which are independently normally distributed with mean 0 and variance σ_e^2.

In the random effect model where the machines under study are considered as a random sample from a large number of machine as also the days are the randomly chosen ones, a_i, b_i, and c_{ij} are all assumed to be independently normally distributed with 0 means and respective variances σ_A^2, σ_B^2 and σ_{AB}^2.

In the fixed effects model where the conclusions are to be drawn only on the few machines that are under study and on the specific days chosen in the experiment, it is assumed that

$$\sum_i a_i = \sum_j b_j = \sum_i c_{ij} = \sum_j c_{ij} = 0$$

In a mixed effect model where the machines are considered as a random sample from a large number of machines, but the days are those specifically chosen, a_i and c_{ij} are assumed to be independently normally distributed with 0 mean and variances σ_A^2 and σ_{AB}^2 and $\sum_j b_j = \sum_j c_{ij} = 0$.

BUREAU OF INDIAN STANDARDS

Headquarters :

Manak Bhavan, 9 Bahadur Shah Zafar Marg, NEW DELHI 110002

Telephones : 331 01 31, 331 13 75 | Telegrams : Manaksanstha (Common to all Offices)

Regional Offices :		*Telephone*
Central	: Manak Bhavan, 9, Bahadur Shah Zafar Marg, NEW DELHI 110002	331 01 31 331 13 75
* Eastern	: 1/14 C.I.T. Scheme VII M, V.I.P. Road, Maniktola, CALCUTTA 700054	37 86 62
Northern	: SCO 445-446, Sector 35-C, CHANDIGARH 160036	2 18 43
Southern	: C.I.T. Campus, IV Cross Road, MADRAS 600113	41 29 16
† Western	: Manakalaya, E9 MIDC, Marol, Andheri (East), BOMBAY 400093	6 32 92 95

Branch Offices :	
'Pushpak', Nurmohamed Shaikh Marg, Khanpur, AHMADABAD 380001	2 63 48
‡ Peenya Industrial Area, 1st Stage, Bangalore-Tumkur Road, BANGALORE 560058	39 49 55
Gangotri Complex, 5th Floor, Bhadbhada Road, T.T. Nagar, BHOPAL 462003	55 40 21
Plot No. 82/83, Lewis Road, BHUBANESHWAR 751002	5 36 27
Kalai Kathir Building, 6/48-A Avanasi Road, COIMBATORE 641037	2 67 05
Quality Marking Centre, N.H. IV, N.I.T., FARIDABAD 121001	—
Savitri Complex, 116 G. T. Road, GHAZIABAD 201001	8-71 19 96
53/5 Ward No. 29, R.G. Barua Road, 5th By-lane, GUWAHATI 781003	3 31 77
5-8-56C L. N. Gupta Marg, (Nampally Station Road) HYDERABAD 500001	23 10 83
R14 Yudhister Marg, C Scheme, JAIPUR 302005	6 34 71
117/418 B Sarvodaya Nagar, KANPUR 208005	21 68 76
Plot No. A-9, House No. 561/63, Sindhu Nagar, Kanpur Road, LUCKNOW 226005	5 55 07
Patliputra Industrial Estate, PATNA 800013	6 23 05
District Industries Centre Complex, Bagh-e-Ali Maidan, SRINAGAR 190011	—
T. C. No. 14/1421, University P. O., Palayam, THIRUVANANTHAPURAM 695034	6 21 04

Inspection Offices (With Sale Point) :	
Pushpanjali, First Floor, 205-A West High Court Road, Shankar Nagar Square, NAGPUR 440010	52 51 71
Institution of Engineers (India) Building, 1332 Shivaji Nagar, PUNE 411005	5 24 35

---

*Sales Office Calcutta is at 5 Chowringhee Approach, P. O. Princep Street, CALCUTTA	27 68 00
† Sales Office is at Novelty Chambers, Grant Road, BOMBAY	89 65 28
‡ Sales Office is at Unity Building, Narasimharaja Square, BANGALORE	22 39 71

Reprography Unit, BIS, New Delhi, India